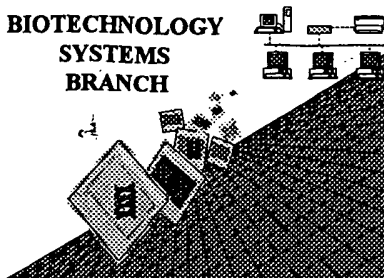# RAW SEQUENCE LISTING ERROR REPORT

BIOTECHNOLOGY
SYSTEMS
BRANCH

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: _09/943,689_

Source: _OIPE_

Date Processed by STIC: _9/20/2001_

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.
PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:
1) **INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY** or,
2) **TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY**
FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.
PATENTIN 2.1 e-mail help: patin21help@uspto.gov **or phone 703-306-4119 (R. Wax)**
PATENTIN 3.0 e-mail help: patin3help@uspto.gov **or phone 703-306-4119 (R. Wax)**

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:

Checker Version 3.0

The Checker Version 3.0 application is a state-of the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.
Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO). Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the f llowing address:
http://www.uspto.gov/web/offices/pac/checker

RAW SEQUENCE LISTING   DATE: 09/20/2001
PATENT APPLICATION: US/09/943,689   TIME: 13:29:07

Input Set : **A:\Seqlist.txt**
Output Set: **N:\CRF3\09202001\I943689.raw**

Does Not Comply
Corrected Diskette Needed

```
 4  <110>  APPLICANT: David Michalovich
 5          Matthew Alan Sims
 6          Narjis Shaikh
 8  <120>  TITLE OF INVENTION: NOVEL COMPOUNDS
11  <130>  FILE REFERENCE: GP-30088-D1
13  <140>  CURRENT APPLICATION NUMBER: US/09/943,689
14  <141>  CURRENT FILING DATE: 2001-08-31
16  <150>  PRIOR APPLICATION NUMBER: US 09/184,001
17  <151>  PRIOR FILING DATE: 1998-11-02
19  <150>  PRIOR APPLICATION NUMBER: UK 9806221.9
20  <151>  PRIOR FILING DATE: 1998-03-23
22  <150>  PRIOR APPLICATION NUMBER: UK 9817479.0
23  <151>  PRIOR FILING DATE: 1998-08-11
25  <160>  NUMBER OF SEQ ID NOS: 4
27  <170>  SOFTWARE: FastSEQ for Windows Version 3.0
29  <210>  SEQ ID NO: 1
30  <211>  LENGTH: 2186
31  <212>  TYPE: DNA
32  <213>  ORGANISM: HOMO SAPIENS
34  <400>  SEQUENCE: 1
35   cttggcacga ggcctcgtgc caggcggcat gagggcccgc ggcccggggg gctgaggcgc    60
36   ccgccgcctg ccgcggggggc cgctcgcgtc ctccatggag gccggagagg aaccgctgct   120
37   gctggccgaa ctcaagcccg ggcgccccca ccagtttgat tggaagtcca gctgtgaaac   180
38   ctggagcgtc gccttctccc cagatggctc ctggtttgct tggtctcaag gacactgcat   240
39   cgtcaaactg atcccctggc cgttggagga gcagttcatc cctaaagggt ttgaagccaa   300
40   aagccgaagt agcaaaaatg agacgaaagg gcggggcagc ccaaaagaga agacgctgga   360
41   ctgtggtcag attgtctggg ggctggcctt cagcccgtgg ccttccccac ccagcaggaa   420
42   gctctgggca cgccaccacc cccaagtgcc cgatgtctct tgcctggttc ttgctacggg   480
43   actcaacgat gggcagatca agatctggga ggtgcagaca gggctcctgc ttttgaatct   540
44   ttccggccac caagatgtcg tgagagatct gagcttcaca cccagtggca gtttgatttt   600
45   ggtctccgcg tcacgggata agactcttcg catctgggac ctgaataaac acggtaaaca   660
46   gattcaagtg ttatcgggcc acctgcagtg ggtttactgc tgttccatct ccccagactg   720
47   cagcatgctg tgctctgcag ctggagagaa gtcggtcttt ctatggagca tgaggtccta   780
48   cacgttaatt cggaagctag agggccatca aagcagtgtt gtctcttgtg acttctcccc   840
49   cgactctgcc ctgcttgtca cggcttctta cgataccaat gtgattatgt gggacccta   900
50   caccggcgaa aggctgaggt cactccacca cacccaggtt gaccccgcta tggatgacag   960
51   tgacgtccac attagctcac tgagatctgt gtgcttctct ccagaaggct tgtaccttgc  1020
52   cacggtggca gatgacagac tcctcaggat ctgggccctg gaactgaaaa ctcccattgc  1080
53   atttgctcct atgaccaatg ggctttgctg cacatttttt ccacatggtg gagtcattgc  1140
54   cacagggaca agagatggcc acgtccagtt ctggacagct cctagggtcc tgtcctcact  1200
55   gaagcactta tgccggaaag cccttcgaag tttcctaaca acttaccaag tcctagcact  1260
56   gccaatcccc aagaaaatga aagagttcct cacatacagg acttttaag caacaccaca  1320
57   tcttgtgctt ctttgtagca gggtaaatcg tcctgtcaaa gggagttgct ggaataatgg  1380
58   gccaaacatc tggtcttgca ttgaaatagc atttctttgg gattgtgaat agaatgtagc  1440
59   aaaaccagat tccagtgtac tagtcatgga tctttctctc cctggcatgt gaaagtcagt  1500
60   cttagaggaa gagattccac ttgcacggca acagagcctt acgttaaatt ttcagtccag  1560
61   ttatgaacag caagtgttga actctttctg cttgttttga ttcaaagtgc agttactgat  1620
```

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/943,689

DATE: 09/20/2001
TIME: 13:29:07

Input Set : A:\Seqlist.txt
Output Set: N:\CRF3\09202001\I943689.raw

```
62  gttgttttga ttatgcaact aagtaggcct ccagagcctc tctagtggca gagcagctca   1680
63  cactccctcc gctgggaacg atggcttctg cctagtacct atccttgtgt ttctgatgca   1740
64  gtggtagcat tggttcaagt tctctcctgc tgtggtcaga gttgcttcga tgttggccaa   1800
65  gtgctttttct tcttgggctc ccttctgacc tgcaggacag ttttcctgga gccatttggt   1860
66  atgaggtatt aatttagctt aactaaatta caggggactc agaggccgtg ctcctgaccg   1920
67  atccagacac tattactggc ttttttttttt tttttttaa caatggtgtg catgtgcagg   1980
68  aaatgacaaa tttgtatgtc agattataca aggatgtatt cttaaaccgc atgactattc   2040
69  agatggctac tgagttatca gtggccattt attagcatca tatttatttg tattttctca   2100
70  acagatgtta aggtacaact gtgttttct cgattatcta aaaccatag tacttaaatt   2160
71  gaaaaaaaaa aaaaaaaaaa aaaaa                                          2186
73  <210> SEQ ID NO: 2
74  <211> LENGTH: 404
75  <212> TYPE: PRT
76  <213> ORGANISM: HOMO SAPIENS
78  <400> SEQUENCE: 2
79  Met Glu Ala Gly Glu Glu Pro Leu Leu Leu Ala Glu Leu Lys Pro Gly
80  1               5                  10                  15
81  Arg Pro His Gln Phe Asp Trp Lys Ser Ser Cys Glu Thr Trp Ser Val
82              20                  25                  30
83  Ala Phe Ser Pro Asp Gly Ser Trp Phe Ala Trp Ser Gln Gly His Cys
84          35                  40                  45
85  Ile Val Lys Leu Ile Pro Trp Pro Leu Glu Glu Gln Phe Ile Pro Lys
86          50                  55                  60
87  Gly Phe Glu Ala Lys Ser Arg Ser Ser Lys Asn Glu Thr Lys Gly Arg
88  65                  70                  75                  80
89  Gly Ser Pro Lys Glu Lys Thr Leu Asp Cys Gly Gln Ile Val Trp Gly
90                  85                  90                  95
91  Leu Ala Phe Ser Pro Trp Pro Ser Pro Pro Ser Arg Lys Leu Trp Ala
92              100                 105                 110
93  Arg His His Pro Gln Val Pro Asp Val Ser Cys Leu Val Leu Ala Thr
94              115                 120                 125
95  Gly Leu Asn Asp Gly Gln Ile Lys Ile Trp Glu Val Gln Thr Gly Leu
96          130                 135                 140
97  Leu Leu Leu Asn Leu Ser Gly His Gln Asp Val Val Arg Asp Leu Ser
98  145                 150                 155                 160
99  Phe Thr Pro Ser Gly Ser Leu Ile Leu Val Ser Ala Ser Arg Asp Lys
100                 165                 170                 175
101 Thr Leu Arg Ile Trp Asp Leu Asn Lys His Gly Lys Gln Ile Gln Val
102             180                 185                 190
103 Leu Ser Gly His Leu Gln Trp Val Tyr Cys Cys Ser Ile Ser Pro Asp
104             195                 200                 205
105 Cys Ser Met Leu Cys Ser Ala Ala Gly Glu Lys Ser Val Phe Leu Trp
106         210                 215                 220
107 Ser Met Arg Ser Tyr Thr Leu Ile Arg Lys Leu Glu Gly His Gln Ser
108 225                 230                 235                 240
109 Ser Val Val Ser Cys Asp Phe Ser Pro Asp Ser Ala Leu Leu Val Thr
110                 245                 250                 255
111 Ala Ser Tyr Asp Thr Asn Val Ile Met Trp Asp Pro Tyr Thr Gly Glu
112             260                 265                 270
```

RAW SEQUENCE LISTING   DATE: 09/20/2001
PATENT APPLICATION: US/09/943,689   TIME: 13:29:07

Input Set : A:\Seqlist.txt
Output Set: N:\CRF3\09202001\I943689.raw

```
113  Arg Leu Arg Ser Leu His His Thr Gln Val Asp Pro Ala Met Asp Asp
114          275             280             285
115  Ser Asp Val His Ile Ser Ser Leu Arg Ser Val Cys Phe Ser Pro Glu
116      290             295             300
117  Gly Leu Tyr Leu Ala Thr Val Ala Asp Asp Arg Leu Leu Arg Ile Trp
118  305             310             315             320
119  Ala Leu Glu Leu Lys Thr Pro Ile Ala Phe Ala Pro Met Thr Asn Gly
120              325             330             335
121  Leu Cys Cys Thr Phe Phe Pro His Gly Gly Val Ile Ala Thr Gly Thr
122          340             345             350
123  Arg Asp Gly His Val Gln Phe Trp Thr Ala Pro Arg Val Leu Ser Ser
124          355             360             365
125  Leu Lys His Leu Cys Arg Lys Ala Leu Arg Ser Phe Leu Thr Thr Tyr
126      370             375             380
127  Gln Val Leu Ala Leu Pro Ile Pro Lys Lys Met Lys Glu Phe Leu Thr
128  385             390             395             400
129  Tyr Arg Thr Phe
132  <210> SEQ ID NO: 3
133  <211> LENGTH: 2558
134  <212> TYPE: DNA
135  <213> ORGANISM: HOMO SAPIENS
136  <220> FEATURE: FEATURE
136  <220> FEATURE: FEATURE
137  <221> NAME/KEY: UNSURE
138  <222> LOCATION: (146)(161)(178)(233)(254)(296)
139  <223> OTHER INFORMATION: OTHER INFORMATION: Partial DNA sequence derived from EST
sequences.
141  <400> SEQUENCE: 3
142  gaaaggaacc gctgctgctg gccgaactca agcccgggcg cccccaccag tttgattgga    60
143  agtccagctg tgaaacctgg agcgtcgcct tctccccaga tggctcctgg tttgcttggt   120
144  ctcaaggaca_ctgcatcgtc aaactnatcc cctggccgtt ngaggagcag ttcatccnta   180
145  aagggtttga agccaaaagc cgaagtagca aaaatgagac gaaagggcgg ggnagtccaa   240
146  aagagaagac gctngactgt ggtcagattg tctgggggct ggccttcagc ctgtgncttt   300
147  ccccacccag caggaagctc tgggcacgcc accaccccca agtgcccgat gtctcttgcc   360
148  tggttcttgc tacgggactc cacgatgggc agatcaagat ctgggaggtg cagacagggc   420
149  tcctgctttt gaatctttcc ggccaccaag atgtcgtgag agatctgagc ttcacaccca   480
150  gtggcagttt gattttggtc tccgcgtcac gggataagac tcttcgcatc tgggacctga   540
151  ataaacacgg taaacagatt caagtgttat cgggccacct gcagtgggtt tactgctgtt   600
152  ccatctcccc agactgcagc atgctgtgct ctgcagctgg agagaagtcg gtctttctat   660
153  ggagcatgag gtcctacacg ttaattcgga agctagaggg ccatcaaagc agtgttgtct   720
154  cttgtgactt ctcccccgac tctgccctgc ttgtcacggc ttcttacgat accaatgtga   780
155  ttatgtggga cccctacacc ggcgaaaggc tgaggtcact ccaccacacc caggttgacc   840
156  ccgccatgga tgacagtgac gtccacatta gctcactgag atctgtgtgc ttctctccag   900
157  aaggcttgta ccttgccacg gtggcagatg acagactcct caggatctgg gccctggaac   960
158  tgaaaactcc cattgcattt gctcctatga ccaatgggct ttgctgcaca tttttttccac  1020
159  atggtggagt cattgccaca gggacaagag atggccacgt ccagttctgg acagctccta  1080
160  gggtcctgtc ctcactgaag cacttatgcc ggaaagccct tcgaagtttc ctaacaactt  1140
161  accaagtcct agcactgcca atccccaaga aaatgaaaga gttcctcaca tacaggactt  1200
162  tttaagcaac accacatctt gtgcttcttt gtagcagggt aaatcgtcct gtcaaaggga  1260
163  gttgctggaa taatgggcca aacatctggt cttgcattga aatagcattt ctttgggatt  1320
```

[Handwritten annotations in margins:]

W--> 136 do not insert an alphabetical heading on
W--> 136 (same line as numeric)

<2217>, <2220>, and <2223> are required for n's and Xaa's identifier

delete

W--> 144
W--> 145
W--> 146

The CRF program inserts alphabetical headings for clarity.

```
164   gtgaatagaa tgtagcaaaa ccagattcca gtgtactagt catggatctt tctctccctg    1380
165   gcatgtgaaa gtcagtctta gaggaagaga ttccacttgc acggcaacag agccttacgt    1440
166   taaattttca gtccagttat gaacagcaag tgttgaactc tttctgcttg ttttgattca    1500
167   aagtgcagtt actgatgttg ttttgattat gcaactaagt aggcctccag agcctctcta    1560
168   gtggcagagc agctcacact ccctccgctg ggaacgatgg cttctgccta gtacctatcc    1620
169   ttgtgtttct gatgcagtgg tagcattggt tcaagttctc tcctgctgtg gtcagagttg    1680
170   cttcgatgtt ggccaagtgc ttttcttctt gggctccctt ctgacctgca ggacagtttt    1740
171   cctggagcca tttggtatga ggtattaatt tagcttaact aaattacagg ggactcagag    1800
172   gccgtgctcc tgaccgatcc agacactatt actggctttt tttttttttt tttaacaatg    1860
173   gtgtgcatgt gcaggaaatg acaaatttgt atgtcagatt atacaaggat gtattcttaa    1920
174   accgcatgac tattcagatg gctactgagt tatcagtggc catttattag catcatattt    1980
175   atttgtattt tctcaacaga tgttaaggta caactgtgtt tttctcgatt atctaaaaac    2040
176   catagtactt aaattgaaca gttgcaaaga tgtcttaatt gtgtaaagaa ttggtgtagt    2100
177   catgacttta gctgatactc ttatgtacga gatctgtctc tgctgtttaa cttcattgga    2160
178   ttaatcagct ggtttcaact ctactgcgaa acaaaaatag ctccttaaaa gtactgttct    2220
179   ccttcagtgg catgtagtta tctaatcaag acacctcatt caaacaaaac ctgccttagg    2280
180   aaaatttaat atattttaaa ttattttaaa agaaatacaa catcttattc tttagctttc    2340
181   ttaatcggtg ctttatggag gccagtgtaa cgttacatga ctcgttgaga aagttgagga    2400
182   atttcctcta ccacctttgt tgcttgaaga aaaacatgtc ttttcaaaat gagaggcttt    2460
183   cattgaagaa aagaaaaaaa caacagttaa aagctattgg ctctctgttt catttttttc    2520
184   cattaagaaa aaaaaaagtc cccttttaaa acaagcag                            2558
186   <210> SEQ ID NO: 4
187   <211> LENGTH: 400
188   <212> TYPE: PRT
189   <213> ORGANISM: HOMO SAPIENS
```

W--> 190   <220> FEATURE: ~~FEATURE~~,
W--> 190   ~~<220> FEATURE: FEATURE~~.

```
191   <221> NAME/KEY: UNSURE
192   <222> LOCATION: (53)(59)(98)
```

This is a peptide sequence

```
193   <223> OTHER INFORMATION: OTHER INFORMATION: Partial DNA sequence derived from EST
      sequences.
195   <400> SEQUENCE: 4
196   Lys Glu Pro Leu Leu Leu Ala Glu Leu Lys Pro Gly Arg Pro His Gln
197   1               5                   10                  15
198   Phe Asp Trp Lys Ser Ser Cys Glu Thr Trp Ser Val Ala Phe Ser Pro
199               20                  25                  30
200   Asp Gly Ser Trp Phe Ala Trp Ser Gln Gly His Cys Ile Val Lys Leu
201           35              40                  45
```

(W)K-> 202   **Ile Pro Trp Pro Xaa Glu Glu Gln Phe Ile Xaa Lys Gly Phe Glu Ala**

```
203           50              55                  60
204   Lys Ser Arg Ser Ser Lys Asn Glu Thr Lys Gly Arg Gly Ser Pro Lys
205   65                  70                  75                  80
206   Glu Lys Thr Leu Asp Cys Gly Gln Ile Val Trp Gly Leu Ala Phe Ser
207                   85                  90                  95
```

W-> 208   **Leu Xaa Leu Ser Pro Pro Ser Arg Lys Leu Trp Ala Arg His His Pro**

```
209               100                 105                 110
210   Gln Val Pro Asp Val Ser Cys Leu Val Leu Ala Thr Gly Leu His Asp
211           115                 120                 125
212   Gly Gln Ile Lys Ile Trp Glu Val Gln Thr Gly Leu Leu Leu Leu Asn
213       130                 135                 140
```

RAW SEQUENCE LISTING   DATE: 09/20/2001
PATENT APPLICATION: US/09/943,689   TIME: 13:29:07

Input Set : A:\Seqlist.txt
Output Set: N:\CRF3\09202001\I943689.raw

```
214  Leu Ser Gly His Gln Asp Val Val Arg Asp Leu Ser Phe Thr Pro Ser
215  145              150              155              160
216  Gly Ser Leu Ile Leu Val Ser Ala Ser Arg Asp Lys Thr Leu Arg Ile
217              165              170              175
218  Trp Asp Leu Asn Lys His Gly Lys Gln Ile Gln Val Leu Ser Gly His
219              180              185              190
220  Leu Gln Trp Val Tyr Cys Cys Ser Ile Ser Pro Asp Cys Ser Met Leu
221              195              200              205
222  Cys Ser Ala Ala Gly Glu Lys Ser Val Phe Leu Trp Ser Met Arg Ser
223      210              215              220
224  Tyr Thr Leu Ile Arg Lys Leu Glu Gly His Gln Ser Ser Val Val Ser
225  225              230              235              240
226  Cys Asp Phe Ser Pro Asp Ser Ala Leu Leu Val Thr Ala Ser Tyr Asp
227              245              250              255
228  Thr Asn Val Ile Met Trp Asp Pro Tyr Thr Gly Glu Arg Leu Arg Ser
229              260              265              270
230  Leu His His Thr Gln Val Asp Pro Ala Met Asp Asp Ser Asp Val His
231          275              280              285
232  Ile Ser Ser Leu Arg Ser Val Cys Phe Ser Pro Glu Gly Leu Tyr Leu
233          290              295              300
234  Ala Thr Val Ala Asp Asp Arg Leu Leu Arg Ile Trp Ala Leu Glu Leu
235  305              310              315              320
236  Lys Thr Pro Ile Ala Phe Ala Pro Met Thr Asn Gly Leu Cys Cys Thr
237              325              330              335
238  Phe Phe Pro His Gly Gly Val Ile Ala Thr Gly Thr Arg Asp Gly His
239              340              345              350
240  Val Gln Phe Trp Thr Ala Pro Arg Val Leu Ser Ser Leu Lys His Leu
241          355              360              365
242  Cys Arg Lys Ala Leu Arg Ser Phe Leu Thr Thr Tyr Gln Val Leu Ala
243      370              375              380
244  Leu Pro Ile Pro Lys Lys Met Lys Glu Phe Leu Thr Tyr Arg Thr Phe
245  385              390              395              400
```

VERIFICATION SUMMARY   DATE: 09/20/2001
PATENT APPLICATION: **US/09/943,689**   TIME: 13:29:08

Input Set : **A:\Seqlist.txt**
Output Set: **N:\CRF3\09202001\I943689.raw**

L:13 M:270 C: Current Application Number differs, Replaced Current Application Number
L:14 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:136 M:283 W: Missing Blank Line separator, <220> field identifier
L:136 M:256 W: Invalid Numeric Header Field, <220> has non-blank data
L:144 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:145 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:146 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:190 M:283 W: Missing Blank Line separator, <220> field identifier
L:190 M:256 W: Invalid Numeric Header Field, <220> has non-blank data
L:202 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4
L:208 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4